university of groningen

University Medical Center Groningen

University of Groningen

Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury

Fuermaier, Anselm B M; Tucha, Oliver; Russ, Dorothea; Ehrenstein, Johanna K; Stanke, Martin; Heindorf, Ralf; Buggenthin, Rieka; Aschenbrenner, Steffen; Koerts, Janneke; Tucha, Lara

Published in:
Journal of Clinical and Experimental Neuropsychology

DOI:
10.1080/13803395.2019.1710468

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Fuermaier, A. B. M., Tucha, O., Russ, D., Ehrenstein, J. K., Stanke, M., Heindorf, R., Buggenthin, R., Aschenbrenner, S., Koerts, J., & Tucha, L. (2020). Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury. *Journal of Clinical and Experimental Neuropsychology*, *42*(3), 285-297. https://doi.org/10.1080/13803395.2019.1710468

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Routledge
Taylor & Francis Group

Journal of Clinical and Experimental Neuropsychology

ISSN: 1380-3395 (Print) 1744-411X (Online) Journal homepage: https://www.tandfonline.com/loi/ncen20

Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury

Anselm B. M. Fuermaier, Oliver Tucha, Dorothea Russ, Johanna K. Ehrenstein, Martin Stanke, Ralf Heindorf, Rieka Buggenthin, Steffen Aschenbrenner, Janneke Koerts & Lara Tucha

To cite this article: Anselm B. M. Fuermaier, Oliver Tucha, Dorothea Russ, Johanna K. Ehrenstein, Martin Stanke, Ralf Heindorf, Rieka Buggenthin, Steffen Aschenbrenner, Janneke Koerts & Lara Tucha (2020) Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury, Journal of Clinical and Experimental Neuropsychology, 42:3, 285-297, DOI: 10.1080/13803395.2019.1710468

To link to this article: https://doi.org/10.1080/13803395.2019.1710468

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 10 Feb 2020.

Submit your article to this journal

Article views: 574

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ncen20

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2020, VOL. 42, NO. 3, 285–297
https://doi.org/10.1080/13803395.2019.1710468

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Utility of an attention-based performance validity test for the detection of feigned cognitive dysfunction after acquired brain injury

Anselm B. M. Fuermaier[a], Oliver Tucha[a], Dorothea Russ[a], Johanna K. Ehrenstein[a], Martin Stanke[b], Ralf Heindorf[b], Rieka Buggenthin[c], Steffen Aschenbrenner[d], Janneke Koerts[a] and Lara Tucha[a]

[a]Department of Clinical and Developmental Neuropsychology, Faculty of Behavioural and Social Science, University of Groningen, Groningen, The Netherlands; [b]Neurological Rehabilitation Friedehorst, Bremen, Germany; [c]Department of Neurology, Marcus Hospital Bad Driburg, Bad Driburg, Germany; [d]Department of Clinical Psychology and Neuropsychology, SRH Clinic Karlsbad-Langensteinbach, Karlsbad, Germany

ABSTRACT

Introduction: The Groningen Effort Test (GET) is a recently developed performance validity test (PVT) for the identification of noncredible performance in a neuropsychological assessment of attention abilities. Because the majority of already established PVTs are based on memory functions, the GET has the potential to make a valuable contribution to validity testing.

Method: The current study examined the utility of the GET in the detection of feigned cognitive dysfunction after acquired brain injury (ABI) and its incremental validity over already established PVTs, namely the Test of Memory Malingering (TOMM), the Dot Counting Test (DCT), and the b Test. Three hundred and forty-eight participants took part in this study, including 58 patients with ABI (stroke or traumatic brain injury), 43 healthy individuals instructed to show normal behavior, and 247 healthy individuals instructed to feign cognitive dysfunction after ABI.

Results: With excellent overall classification accuracy, the GET performed close to the level of the TOMM, and superior to the b Test and DCT. Data analyses further revealed that the GET provides additional diagnostic accuracy compared to the b Test and the DCT in the detection of feigned cognitive dysfunction, but has no incremental validity over the TOMM. For each of the four PVTs in this study, diagnostic sensitivity was independent of the simulation strategy used.

Conclusions: It is concluded that the GET is an attention-based PVT with promising test characteristics and high diagnostic accuracy in the detection of noncredible cognitive performance using a simulation design. Given the results can be replicated in studies using *known-groups* methodology, it may be a useful tool for clinical practice to complement neuropsychological assessments of patients with ABI.

ARTICLE HISTORY
Received 24 October 2019
Accepted 24 December 2019

KEYWORDS
Neuropsychological assessment; attention; acquired brain injury; performance validity; feigning

Patients with acquired brain injury (ABI; such as traumatic brain injury or stroke) frequently suffer from a range of cognitive dysfunctions, including difficulties in attention, concentration, memory, perception, and action planning (Jennekens, de Casterlé, & Dobbels, 2010; Sun, Tan, & Yu, 2014). Routine clinical practice therefore often includes a comprehensive neuropsychological assessment of cognitive functions in patients with ABI, in order to objectively measure the cognitive impairments reported by patients, characterize cognitive strengths and weaknesses, and guide treatment planning (Lezak, 2004). In this context, position papers and consensus reports state that validity testing should be an integral part of any routine neuropsychological assessment (Bush et al., 2005; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009; Larrabee, 2007), allowing the clinician to make conclusions about both the integrity of function and credibility of demonstrated symptoms during assessment.

Estimates regarding the prevalence of feigned or exaggerated cognitive dysfunction are difficult to derive and depend on many factors, including the context, the sample studied, and the instruments used to identify feigned or exaggerated cognitive dysfunction (Dandachi-FitzGerald, Ponds, & Merten, 2013; Greiffenstein, Baker, & Gola, 1994; Mittenberg, Patton, Canyock, & Condit, 2002). However, there is broad consensus on the importance of the detection of noncredible symptom reporting and performance in an early phase of clinical care on the basis of instruments designed and validated for this purpose (Bush et al., 2005; Haines & Norris, 2001; Heilbronner et al., 2009).

Various performance validity tests (PVTs) have been introduced, with a considerable number of the most widely used PVTs being memory-based, such as the Test of Memory Malingering (TOMM; Tombaugh, 1996), the Word Memory Test (WMT; Green, Allen, & Astner, 1996), the Medical Symptom Validity Test (MSVT;

CONTACT Anselm B. M. Fuermaier a.b.m.fuermaier@rug.nl Faculty of Behavioural and Social Sciences, Department of Clinical and Developmental Neuropsychology, University of Groningen, Grote Kruisstraat 2/1, Groningen 9712 TS, The Netherlands
Supplemental data for this article can be accessed here.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Green, 2005a), and the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2005b). Other established PVTs are based on different principles, such as visual perception (e.g., Dot Counting Test; DCT; Boone, Lu, & Herzberg, 2002a), or letter recognition and discrimination (e.g., b Test; Boone, Lu, & Herzberg, 2002b). While the psychometric properties of these PVTs have been frequently studied (Binks, Gouvier, & Waters, 1997; Rees, Tombaugh, Gansler, & Moczynski, 1998), and while these tests are certainly useful for clinical practice to detect feigned memory deficits, they might be limited in their accuracy to detect individuals feigning attention deficits, such as distraction, lapses of attention, mind wandering, or problems with sustaining attention. The importance of assessing noncredible attention performance is underlined by qualitative research on the commonly applied deception strategies, which have identified, next to memory loss (76%), slow response time (32%), poor concentration (12%), and confusion (16%) as the most common strategies (Tan, Slick, Strauss, & Hultsch, 2002). Because simulated attention deficits, as constituted by slow response time and poor concentration, make up to 44% of the most commonly used simulation strategies (Tan et al., 2002), it can be concluded that there is a need for validity tests that are specifically based on attention performance.

The Groningen Effort Test (GET; Fuermaier, Tucha, Koerts, Aschenbrenner, & Tucha, 2017; Fuermaier et al., 2016) is an attention-based PVT that has been recently developed to detect the noncredible attention performance of individuals during psychiatric evaluation. In the context of adult attention-deficit/hyperactivity disorder (ADHD) using a simulation design, the GET shows excellent diagnostic accuracy in the detection of feigned attention deficits, with 88% sensitivity toward feigned ADHD and 90% specificity toward genuine ADHD. Nevertheless, to date, little is known about the GET's utility to differentiate between credible and noncredible cognitive test performance among patients with ABI.

The present study aims to address this issue and examines the utility of the GET in the detection of the noncredible cognitive performance of patients after ABI. This study further aims to determine whether individuals applying different simulation strategies can be better identified with the GET rather than already established PVTs. These objectives are addressed in a study employing a simulation design with five groups: healthy individuals allocated to either a healthy control group required to perform all tests to the best of their abilities, or to one of four simulation groups asked to perform the measures as if they suffered from cognitive dysfunction following ABI. The PVT performances of these groups are compared to the performance of a group of patients with stroke or TBI with genuine cognitive dysfunctions. We expected the GET to be at least as useful as other PVTs in the detection of feigned cognitive dysfunction after ABI. We further expected the GET to have incremental validity over at least some of the already established PVTs. Finally, we expected the GET to be particularly useful in the detection of simulated attention deficits.

Method

Participants

Power issues are usually not a major concern in this type of research, as relatively large effects are required for the detection of feigning, which are revealed with relatively small samples. Rogers (2008), for example, introduced the classification of effects into moderate if Cohen's $d \geq .75$, and large if Cohen's $d \geq 1.25$. Based on a group comparison with a two-tailed test and $\alpha = .05$, revealing a moderate effect ($d = .75$) with a desired power ($1-\beta$) of .85 requires a sample size of 33 participants per group, while a large effect ($d = 1.25$) requires a sample size of only 13 participants per group. However, a reliable estimation of classification accuracy requires both a sufficient number of individuals feigning cognitive dysfunction (to determine sensitivity), and a sufficient number of patients with genuine cognitive dysfunctions (to determine specificity). We therefore aimed to exceed the minimum number of participants as indicated in the power analysis; attempting to reach a group size of at least 50 patients with ABI with genuine cognitive dysfunctions.

Three hundred and forty-eight participants took part in this prospective study. Patients with ABI (P_{ABI}; $n = 58$) consisted of patients with TBI ($n = 28$) or stroke ($n = 30$), and were recruited and assessed at two neurological rehabilitation hospitals in Germany. Healthy individuals ($n = 292$) included both controls (*CG;* $n = 43$) and instructed simulators (*SIM;* $n = 247$), and were recruited and assessed by trained graduate students.

Patients with acquired brain injury

All patients with ABI were invited to take part in the study on a voluntary basis and were not rewarded for their participation. Patients with ABI included 23 female and 35 male participants ranging from 17 to 77 years (M ± SD = 42.1 ± 16.1). The duration of illness varied between one and 77 months (M ± SD = 8.5 ± 14.4). At the time of assessment, most patients with ABI were on sick leave ($n = 43$), others were employed ($n = 7$), retired ($n = 4$), in training ($n = 2$), or unemployed ($n = 2$). Neuroimaging support of the brain injury was provided for 47 patients. Forty-two patients were on medication at the time of the assessment

(i.e., antidepressants/stimulants, analgesics, or beta blockers). The degree of severity for TBI was assessed using the Glasgow Coma Scale (GCS) and varied between a minor (n = 2), moderate (n = 2), and severe (n = 18) degree of severity. A GCS score was not assessed for six patients with TBI. Patients with stroke either had an ischemic (n = 22) or hemorrhagic (n = 8) stroke with the lesion affecting the left (n = 14), right (n = 12), or both (n = 1) hemispheres (unknown in n = 3). Rehabilitation phases ranged from A (acute treatment) to F (activating long-term care), with the majority of patients being in rehabilitation phase D (n = 50) at the time of assessment. In this phase, patients have successfully completed early mobilization training and are no longer dependent on essential nursing care. One patient was in rehabilitation phase C, meaning that the patient could actively participate in therapy, but was still dependent on curative medicine and nursing care. Three patients were in rehabilitation phase E, which focuses on domestic and vocational rehabilitation. The rehabilitation phase was not provided for four patients.

Eight patients with ABI had to be excluded due to one of the following reasons. Six Patients with ABI failed to show credibility on the TOMM (n = 1) or DCT (n = 4) as measures of performance validity, or did not complete any of the two (n = 1), and two patients with ABI did not perform the GET (see material section for a description of the validity measures). Table 1 presents the characteristics of 50 patients with ABI that were retained in the present study, including the frequency of patients having cognitive impairments (percentile ≤ 10) as indicated by test norms of standard measures of cognition (see material section for a description of measures). Between almost 20% and 50% of patients with TBI performed in the impaired range on each test of cognition. With regard to patients post-stroke, the range regarding the proportion of individuals with particular impairments was 15% (alertness) and 64% (reaction time in selective attention; see Table 1).

Healthy individuals

Instructed simulators and non-simulating healthy controls were recruited from a first-year psychology program and were awarded course credit in exchange for participation. Healthy individuals also had the chance of winning a tablet PC by participating in the research. The instructed simulation group included 247 students. Students' ages ranged from 17 to 34 years (M ± SD = 20.4 ± 2.3) and included 146 female and 101 male participants. Instructed simulators were randomly assigned to one of the four conditions: (1) naïve TBI simulation group (NS_{TBI}): n = 66; age (M ± SD) = 21.1 ± 2.9 years, gender

Table 1. Characteristics of patients with acquired brain injury.

	TBI		Stroke		Total
Descriptives					
N	23		27		50
Age (in years)	36.5 ± 15.5		47.6 ± 15.0		42.5 ± 16.1
Gender (female/male)	6/17		14/13		20/30
Education (in years)	14.0 ± 3.9		13.9 ± 3.6		13.9 ± 3.6
Employment status (unemployed/in training/working/on sick leave/retired)	2/2/2/15/2		0/0/4/21/2		2/2/6/36/4
IQ (vocabulary skills)[a]	96.4 ± 9.0		108.5 ± 14.8		103.3 ± 14.0
Neuroimaging[b]	20		27		47
Medication[c] (yes/no)	16/7		23/4		39/11
TBI severity[d] (minor/moderate/severe)	2/1/14		-		-
Stroke type (hemorrhagic/ischemic)	-		6/21		-
Stroke[e] Hemisphere (right/left/both)	-		11/12/1		-
Duration of illness in months	12.5 ± 15.7		2.6 ± 2.6		7.5 ± 12.1
Rehabilitation phase (C/D/E)[f]	1/17/2		0/25/1		1/42/3
Cognitive functioning		**% percentile ≤ 10**		**% percentile ≤ 10**	
Visuoconstructive abilities[g]	34.4 ± 18.6	28.6	26.0 ± 16.0	18.5	29.7 ± 17.6
Alertness RT[h]	367.8 ± 373.7	34.8	259.1 ± 61.4	15.4	310.1 ± 262.7
Selective Attention RT[i]	624 ± 269	42.9	683 ± 259	63.6	655 ± 263
Selective Attention Om[j]	8.3 ± 6.8	19.0	6.4 ± 7.8	17.4	7.3 ± 7.3
Divided Attention Om[k]	3.8 ± 5.0	40.9	2.1 ± 2.1	25.0	2.9 ± 3.9
Word Fluency[l]	16.2 ± 6.1	47.4	17.6 ± 7.4	33.3	17.0 ± 6.8

[a]Multiple Choice Vocabulary Test (MWT-B);
[b]Brain damage was supported by neuroimaging;
[c]Number of patients known to be taking medication at the time of the assessment: Antidepressants, stimulants, analgesics, beta blockers;
[d]Glasgow Coma Scale (GCS) severity classification for patients with TBI;
[e]Side of the lesion;
[f]German classification of rehabilitation phases after ABI;
[g]Block design of the WAIS-IV;
[h]TAP phasic Alertness reaction time in milliseconds;
[i]TAP Visual Scanning reaction time in seconds;
[j]TAP Visual Scanning Omissions;
[k]TAP Divided Attention Omission;
[l]Regensburg Word Fluency Test (S-words). Stroke = All patients with Stroke; TBI = All patients with TBI; Total = All Patients with ABI.

(female/male) = 40/26; (2) coached TBI simulation group (CS_{TBI}): n = 64; age (M ± SD) = 20.5 ± 2.3 years, gender (female/male) = 38/26; (3) naïve stroke simulation group (NS_{Stroke}) n = 58, age (M ± SD) = 20.0 ± 1.9 years, gender (female/male) = 35/23; and (4) coached stroke simulation group (CS_{Stroke}): n = 59; age (M ± SD) = 19.8 ± 1.7 years, gender (female/male) = 33/26.

Fifty-five healthy simulating individuals were excluded for one or more of the following reasons: 26 failed to score higher than six in the Reliable Digit Span, which served as an indicator of suboptimal effort (Greiffenstein et al., 1994; see material section for a description of all measures), 25 reported after the assessment that they had not put in their best effort to simulate cognitive dysfunction (cutoff < 3), and 8 participants reported having a psychiatric illness.

The non-simulating healthy control group (n = 43) included 29 females and 14 males ranging from 17 to 25 years (M ± SD = 20.0 ± 1.7). Healthy individuals were required to perform and pass three different validity tests (i.e., TOMM, DCT, b Test). Three participants in the control group failed the b Test (Boone et al., 2002b) and were therefore excluded. Table 2 provides characteristics of all participants that were retained in the present study.

Materials

Participants were presented with a comprehensive test battery that included an inventory of descriptive and clinical information, tests for cognition and intellectual functions, several validity tests, and a self-rating of their own performance. Tests were administered verbally, in paper and pencil form, or computerized.

Descriptive and clinical information

All participants were asked about descriptive and clinical information including age, gender, highest level of education, total years of education, current state of employment (i.e., student, unemployed, in training, employed, on sick leave, retired), psychiatric conditions, and medication use. Additional information was acquired from patients with ABI regarding their acquired brain injury, including type of brain injury, duration and severity of injury, whether the brain injury was supported by neuroimaging data, and the rehabilitation phase.

Intellectual functions (vocabulary skills)

Intellectual functions (IQ) were measured using the *Multiple Choice Vocabulary Test*(MWT-B; Lehrl, 1995; Lehrl et al., 1995). The MWT-B has been shown to provide a valid measure for premorbid IQ that is relatively insensitive to ABI (Crawford, Parker, & Besson, 1988; Suslow, 2009).

Routine measures of cognition (only presented to patients with ABI)

The subtest *Block Design* from the Wechsler Adult Intelligence Scale (WAIS-IV; Petermann, 2012) was used to assess visuoconstructive abilities and perceptual logic. Alertness, selective attention, and divided attention were assessed with subtests from the computerized *test battery for attention performance* (TAP; Zimmermann & Fimm, 2010), i.e., *Alertness, Visual Scanning*, and *Divided Attention*. The *Regensburg Word Fluency Test* (RWT; Aschenbrenner, Tucha, & Lange, 2000) was used to assess phonemic word fluency.

Performance validity tests

The *Reliable Digit Span* (RDS; Greiffenstein et al., 1994) is an embedded performance validity measure based on the Digit Span Test of the Wechsler Adult Intelligence Scale (WAIS-IV; Petermann, 2012). An RDS score below the cutoff of seven indicates malingered neurocognitive functions or suboptimal effort (Greve et al., 2007; Heinly, Greve, Bianchini, Love, & Brennan, 2005).

The *Groningen Effort Test* (GET; Fuermaier et al., 2016, 2017) is an attention-based validity test. In the clinical evaluation of adults with ADHD, a GET index ≥ −.25 indicates noncredible performance (GET-ADHD index).

The *Test of Memory Malingering* (TOMM; Tombaugh, 1996) is a visual recognition test consisting of two learning trials (Trials 1 and 2) and a retention trial. Noncredible performance is indicated if the number of correct trials is <45 in either Trial 2 or the retention trial.

Table 2. Characteristics of participants: control group, patients with ABI, and simulation groups (Mean±SD).

	CG	P_{ABI}	Sim	NS_{TBI}	CS_{TBI}	NS_{Stroke}	CS_{Stroke}
N	40	50	192	52	48	46	46
Age (in years)	19.8 ± 1.5	42.5 ± 16.1	20.2 ± 2.2	20.8 ± 2.7	20.3 ± 2.1	20.1 ± 2.0	19.7 ± 1.7
Gender (female/male)	26/14	20/30	113/79	32/20	30/18	27/19	24/22
Education (in years)	13.2 ± 1.3	13.9 ± 3.6	13.5 ± 1.6	13.6 ± 2.0	13.5 ± 1.3	13.3 ± 1.6	13.3 ± 1.4
IQ (vocabulary skills)[a]	[b]	103.3 ± 14.0	97.2 ± 7.9	96.5 ± 6.7	96.9 ± 7.7	98.0 ± 8.2	97.6 ± 9.4

[a]Multiple Choice Vocabulary Test (MWT-B); CG = Control group; P_{ABI} = Patients with ABI; Sim = All instructed simulation groups collapsed; NS_{TBI} = Naïve simulation TBI group; CS_{TBI} = Symptom-coached simulation TBI group; NS_{Stroke} = Naïve simulation stroke group; CS_{Stroke} = Symptom-coached simulation stroke group;
[b]= Not assessed in the control group.

The *Dot Counting Test* (DCT; Boone et al., 2002a) is a stand-alone validity test based on visual perception and basic counting skills. An E-score ≥20 indicates noncredible performance in the TBI group, whereas an E-score ≥22 indicates noncredible performance for stroke patients (Boone et al., 2002a). Additionally, a revised cutoff (E-score ≥ 13.80) is employed as reported by Roberson et al. (2013).

The *b Test* (Boone et al., 2002b) is a letter recognition and discrimination task designed to detect noncredible performance. Cutoffs for TBI patients (E-score ≥ 90) and stroke patients (E-score ≥ 170) are defined (Boone et al., 2002b). Additionally, a revised cutoff of E-score ≥82 is employed as suggested by McCaul et al. (2018).

Procedure

Assessment of patients with acquired brain injury

All patients with ABI were tested individually and received no reward for participation. Written informed consent was sought from all participants prior to the assessment. It was pointed out to patients that all data collected in the research project would be analyzed anonymously and would not affect clinical assessment and treatment. The assessment took approximately 2.5 h. The examiner was present at all times during the assessment, and instructed patients and recorded the answers. Patients with ABI were assessed with a comprehensive test battery including an inventory of descriptive and clinical information, routine measures of cognition, a measure of intellectual functions (vocabulary skills), and PVTs (GET, TOMM, and DCT). The b Test was not presented to patients with ABI.

Assessment of healthy participants

All healthy participants were tested individually in a quiet laboratory. Written consent was obtained from all participants prior to the assessment. The assessment took approximately 1.5 h. An instructor was present at all times during the assessment, and was responsible for test administration and scoring. At the beginning of the assessment, before the simulation, all participants were asked to provide descriptive information. Furthermore, intellectual functions were assessed. The subsequent assessment procedure differed between participants of the various groups (HC, NS_{TBI}, CS_{TBI}, NS_{Stroke}, and CS_{Stroke}).

Assessment of participants in the control group (HC).

Participants in the healthy control group were asked to perform all tests to the best of their abilities. They completed four performance validity tests (GET, TOMM, DCT, and b Test) and received a debriefing.

Assessment of participants in the simulation group.

Participants in the simulation groups were asked to perform the RDS test. Next, participants were randomly presented with one of the four vignettes describing a scenario in which someone would be motivated to feign cognitive dysfunction. At this point, participants in the coached conditions were also provided with simulation instructions. Participants were then asked to complete the following performance validity tests (GET, TOMM, DCT, b Test) while pretending to suffer from cognitive impairment. Finally, participants were asked to stop simulating cognitive dysfunction, reveal their utilized simulation strategies and were debriefed.

Simulation instructions (only presented to individuals simulating cognitive dysfunction)

The simulation instructions (NS_{TBI}, CS_{TBI}, NS_{STROKE}, CS_{STROKE}) varied with regard to the type of brain injury (TBI or stroke) to be simulated and the level of coaching (naïve or coached). The simulation instructions were adapted from previous research on feigned cognitive dysfunction (Arnett, Hammeke, & Schwartz, 1995; Erdal, 2004; Rose, Hall, Szalda-Petree, & Bach, 1998). The complete instructions can be found in the supplementary file (naïve and coached simulation instructions for participants assigned to the TBI and stroke group). Coached groups were instructed not to be too obvious in feigning cognitive dysfunction, in order to produce a believable performance for someone with genuine brain injury (i.e., successful simulation). All simulating participants were instructed to apply their best efforts in feigning cognitive dysfunction. To increase motivation, participants were informed that the most successful simulation would be awarded a top-of-the-range tablet PC. However, due to ethical reasons, the tablet PC was eventually allocated at random to one of the participants.

Strategy use (only presented to individuals simulating cognitive dysfunction)

Participants were asked to reveal their strategies used during simulation and to rate their performance. A five-point Likert scale (ranging from 1 = *strongly disagree* to 5 = *strongly agree*) was used to assess the question of whether they tried their best to simulate cognitive dysfunction. Participants were also asked to specify the strategies they utilized to simulate cognitive dysfunction (i.e., slowed down responses, inattention, memory problems, and disorganization) and were also given the option to name other strategies used. Finally, it was asked which of the tests the participants suspected to be specifically designed to detect feigned cognitive deficits.

Compliance of ethical standards

The study was performed in accordance with the standards of the latest version of the Helsinki declaration. Ethical approval for the assessment of patients with ABI was obtained from the medical ethical committee of the University of Oldenburg, Germany. Ethical permission for the assessment of healthy individuals was obtained from the ethical committee psychology (ECP) at the university the authors were affiliated to.

Statistical analyses

With regard to GET test performance, the results of all test variables were presented separately for control participants, patients with ABI, and instructed simulators. In addition to descriptive analysis, mean reaction time and total number of errors were statistically compared between groups (independent t-test with Cohen's d as effect size). Next, two indices were calculated for the GET per person, i.e., the GET-ADHD index and GET-ABI index. The algorithm for calculating the GET-ADHD index was adopted from the original study of Fuermaier et al. (2016), who examined the use of the GET in differentiating genuine from feigned cognitive dysfunction in the clinical evaluation of adult ADHD. This index was computed by summing up the eight GET scores, each weighted with their respective regression coefficient to detect feigned ADHD. Furthermore, the GET-ABI index was derived from the current data set by summating weighted GET scores. This time, the weights (i.e., regression coefficients) were derived from a logistic regression analysis that predicts group membership (genuine or feigned ABI) on the basis of the eight GET scores as predictors. The rationale for computing this index score was to determine which combination of GET test variables would be best in terms of differentiating genuine from feigned cognitive dysfunction in the present sample. The accuracy of each of the GET indices in detecting individuals feigning ABI (collapsed group of all simulation conditions n = 192) relative to patients with ABI (n = 50) was explored in receiver operating characteristics (ROC) by calculating the area under the curves (AUC). Diagnostic accuracies (i.e., levels of sensitivity and specificity) of the two GET indices were presented and compared to other PVTs (i.e., sensitivity; specificity toward genuine ABI could not be derived as passing the PVTs was an inclusion criterion). In this analysis, the cutoffs for the TOMM, DCT, b Test, and the GET-ADHD index were taken from their respective test manuals or publications (Boone et al., 2002a, 2002b; Fuermaier et al., 2016; McCaul et al., 2018; Roberson et al., 2013; Tombaugh, 1996). In addition to using the established cutoff for the GET-ADHD index, the proposed cutoff from the ADHD research was adapted to achieve a specificity of 90.0% (Boone, 2007). The same criterion (specificity of at least 90%) was applied to determine a cutoff for the GET-ABI index.

In further analyses, the incremental validity of the GET over the TOMM and DCT was explored in partial ROC analyses (pROC). In these analyses, a linear regression was first computed separately for the TOMM and DCT on the newly computed GET-ABI index. The unstandardized residuals from this regression were then entered into an ROC analysis to identify feigned cognitive dysfunction relative to genuine ABI. Of note, for this analysis patients who had failed the TOMM or DCT, and who were therefore disregarded in the analyses described above (in order to ensure credibility of patients with ABI), were additionally included. Because the purpose of the pROC analysis was to determine the diagnostic overlap between two measures (i.e., TOMM or DCT vs. GET), patients who failed either of the tests had to be included in order to consider both rates of sensitivity and specificity.

Finally, simulation strategy use, and recognition of the various PVTs as validity tests, were analyzed in descriptive statistics by reporting absolute and relative frequencies.

Results

Utility of the GET and other established PVTs

With regard to performance on the GET, reaction times and errors per block, mean reaction times, and the total number of errors are shown in Table 3. Independent sample t-tests revealed significant differences in mean reaction times for patients with ABI compared to controls, t (88) = 3.432, p = .001, d = .71, and instructed simulators, t (240) = 2.308, p = .022, d = .36, with patients having longer mean reaction times than both controls and instructed simulators

Table 3. Groningen Effort Test (GET) performance of control participants, patients with ABI, and simulation groups (Mean ±SD).

Measures	CG (n = 40)	P_{ABI} (n = 50)	Sim (n = 192)
Block 1, RT (sec)	3.8 ± 2.5	6.7 ± 5.1	5.0 ± 2.3
Block 1, Errors	1.1 ± 1.4	2.2 ± 2.6	6.6 ± 3.8
Block 2, RT (sec)	3.7 ± 3.4	5.3 ± 3.5	4.4 ± 2.2
Block 2, Errors	1.2 ± 1.3	2.2 ± 2.5	7.4 ± 4.1
Block 3, RT (sec)	2.0 ± 1.9	3.7 ± 2.7	3.5 ± 2.0
Block 3, Errors	.6 ± .7	1.5 ± 2.3	6.7 ± 4.1
Block 4, RT (sec)	1.7 ± 1.3	3.6 ± 3.0	3.1 ± 1.8
Block 4, Errors	.3 ± .5	1.3 ± 2.0	6.7 ± 4.3
Mean RT[a]	2.8 ± 2.2	4.8 ± 3.2[c]	4.0 ± 1.9[c,d]
Total Errors[b]	3.1 ± 2.3	7.2 ± 7.8[c]	27.3 ± 14.8[c,d]

[a]Mean reaction time in seconds for all four blocks combined;
[b]Number of total Errors over all four blocks. CG = Control group; P_{ABI} = Patients with ABI; Sim = All instructed simulation groups collapsed;
[c]Significantly different from CG at p < .05;
[d]Significantly different from P_{ABI} at p < .05.

(Table 3). A significant difference in mean reaction times was also found for healthy controls compared to simulators, t (230) = 3.464, p = .001, d = .61, with controls having shorter mean reaction times. The total number of errors differed significantly for patients with ABI compared to controls, t (88) = 3.214, p = .002, d = .68, for patients with ABI compared to simulators, t (240) = −9.29, p < .001, d = 1.47, and for controls compared to simulators, t (230) = 10.313, p < .001, d = 1.79, with patients making more mistakes than controls, but fewer than instructed simulators.

The ability of the GET to detect simulated cognitive dysfunction (n = 192) relative to genuine pathology (n = 50) was analyzed by the use of two GET indices. First, the GET-ADHD index was computed by summing up the eight GET scores, each weighted with the regression coefficients as adopted from the original study of Fuermaier et al. (2016). Despite being developed for an ADHD sample, the GET-ADHD index also performed well in the present study with regard to the identification of feigned cognitive dysfunction after ABI, *AUC* = .914, *SE* = .021, CI = .872; .956, and p < .001.

Second, for the calculation of the GET-ABI index, the weights (i.e., regression coefficients) were derived from a logistic regression analysis that predicts group membership (genuine or feigned ABI) on the basis of the eight GET scores as predictors. A significant model was found for the identification of feigned ABI (collapsed group of all simulation conditions, n = 192) relative to patients with ABI (n = 50), with χ^2 (8, n = 242) = 138.11, p < .001, explaining 43.5% of the variance (Cox & Snell^2). The prediction equation was as follows: *GET-ABI index* = −.381 *Block 1 RT* + .238 *Block 1 Errors* – .005 *Block 2 RT* + .048 *Block 2 Errors* + .821 *Block 3 RT* + .41 *Block 3 Errors* – .953 *Block 4 RT* + .356 *Block 4 Errors*. The accuracy of the newly developed GET-ABI index in detecting individuals feigning ABI (collapsed group of all simulation conditions, n = 192) relative to patients with ABI (n = 50) was explored in an ROC analysis. In this analysis, the GET-ABI index's ability to identify simulated cognitive dysfunction was excellent: *AUC* = .949, *SE* = .014, CI = .922; .976, and p < .001. The flow of participants and diagnostic accuracy of the GET-ABI index is depicted in Figure 1; a graphical presentation of the ROC curve is shown in Figure 2.

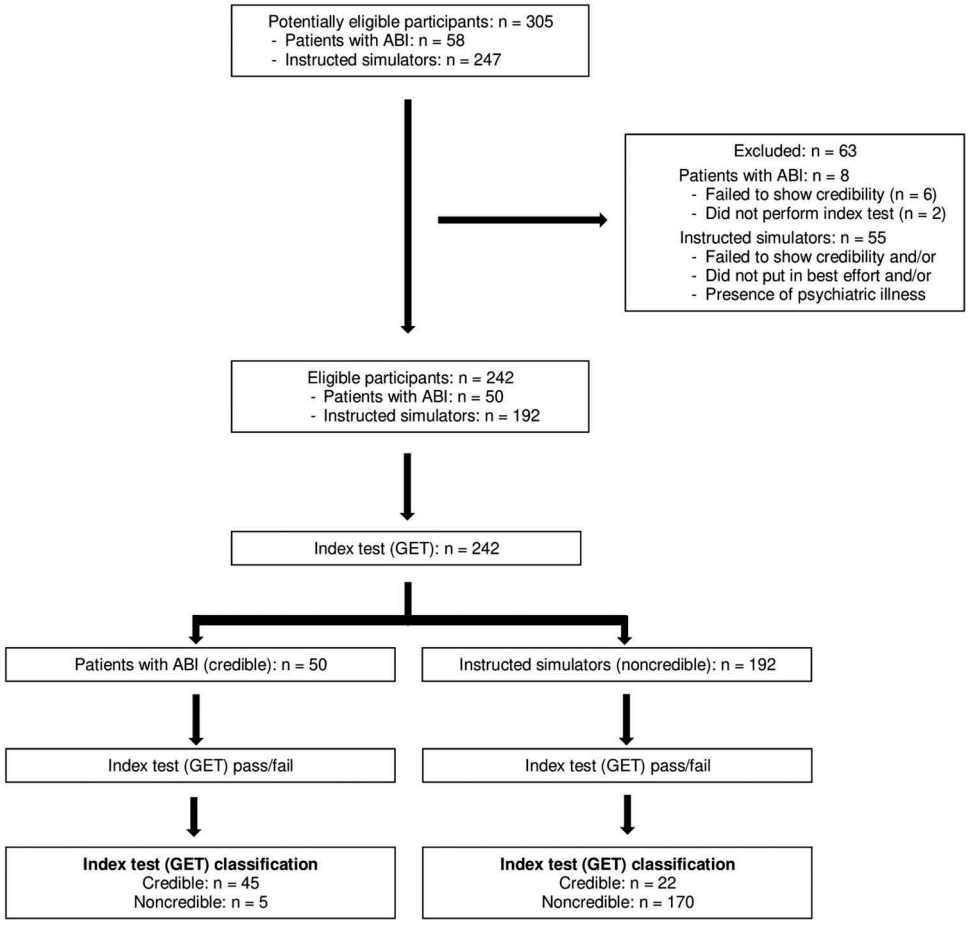

Figure 1. STARD diagram reporting flow of participants and diagnostic accuracy of GET-ABI index (cutoff .791).

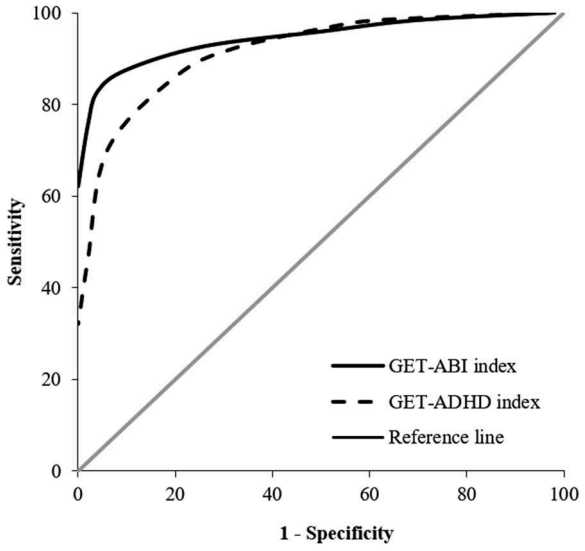

Figure 2. ROC curve depicting diagnostic accuracy of the GET-ABI index and GET-ADHD index in identifying feigned cognitive dysfunction (n = 192) relative to genuine cognitive dysfunction (n = 50).

The test performance of participants on all PVTs, including performance on the TOMM, DCT, b Test and GET, is presented in Table 4. Table 5 presents the ability of validity tests to distinguish feigned from genuine cognitive dysfunction, along with their cutoffs and respective levels of sensitivity, specificity, and AUC. For the GET-ADHD index, the proposed cutoff from ADHD research of ≥ −.25 results in 92.2% sensitivity but only 64.0% specificity. Therefore, an adapted cutoff was determined for the GET-ADHD index to achieve specificity of 90.0% (Boone, 2007), which resulted in a sensitivity level of 82.3%. With regard to the GET-ABI index, achieving specificity of at least 90% resulted in a suitable cutoff of .791 at a sensitivity level of 88.5%. The TOMM's ability to identify feigned cognitive dysfunction was excellent, with a sensitivity of 98.4% at the established cutoff of ≤45 (Tombaugh, 1996). The b Test performed fairly with a sensitivity of 70.3%, at the cutoffs ≥90 and ≥170 for TBI and stroke, respectively (Boone et al., 2002b), but performed well with a sensitivity of 86.5% at the cutoff ≥82 (Roberson et al., 2013). The DCT performed rather poorly with a sensitivity of only 39.1% based on the cutoffs suggested in the test manual (i.e., E-score ≥ 20 for TBI, E-score ≥ 22 for stroke), but performed well with a sensitivity of 80.7% at the cutoff ≥ 13.80 (McCaul et al., 2018).

Table 4. Performance validity tests of control participants, patients with ABI, and simulation groups (Mean ± SD).

	CG (n = 40)	P_{ABI} (n = 50)	Sim (n = 192)
GET-ABI index	−.7 ± .5	−1.3 ± 1.9	5.0 ± 4.0
GET-ADHD index	−1.7 ± 1.7	−1.1 ± 3.1	6.7 ± 5.3
b Test E-score	40.2 ± 11.1	[a]	655.0 ± 1075.0
DCT E-score	8.6 ± 2.0	10.5 ± 3.9	20.4 ± 8.9
TOMM T2	50.0 ± .0	49.9 ± .5	29.5 ± 8.2
TOMM R	50.0 ± .2	49.8 ± .7	28.6 ± 8.3

CG = Control group; P_{ABI} = Patients with ABI; Sim = All instructed simulation groups collapsed; GET = Groningen Effort Test; DCT = Dot Counting Test; TOMM T2 = Test Of Memory Malingering Trial 2; TOMM R = Test Of Memory Malingering Retention Trial;
[a]= Not assessed in patients with ABI.

Table 5. Diagnostic accuracy of various validity tests in distinguishing feigned (n = 192) from genuine (n = 50) cognitive dysfunction, as indicated by levels of sensitivity, specificity, and area under the curve.

Cutoff	Sensitivity	Specificity	AUC
GET-ADHD index			.914
−.25[a]	92.2	64	
2.21[b]	82.3	90	
GET-ABI index			.949
.791[b]	88.5	90	
TOMM			
≤45[c]	98.4	[h]	
DCT (E-Score)			
TBI ≥ 20, Stroke ≥ 22[d]	39.1	[h]	
≥13.80[e]	80.7	[h]	
b Test (E-Score)			
TBI ≥ 90, Stroke ≥ 170[f]	70.3	[i]	
≥82[g]	86.5	[i]	

AUC = Area under the curve; GET = Groningen Effort Test; DCT = Dot Counting Test; TOMM = Test Of Memory Malingering;
[a]= cutoff adopted from ADHD research (Fuermaier et al., 2016);
[b]= cutoff derived from this study in order to maximize specificity to at least 90%;
[c]= cutoff adopted from test manual suggesting noncredible performance: trial 2 or retention trial ≤45 (Tombaugh, 1996);
[d]= cutoff adopted from test manual suggesting noncredible performance: TBI ≥ 20, Stroke ≥ 22 (Boone et al., 2002a);
[e]cutoff adopted from McCaul et al. (2018);
[f]= cutoff adopted from test manual suggesting noncredible performance: TBI ≥ 90, Stroke ≥ 170 (Boone et al., 2002b);
[g]cutoff adopted from Roberson et al. (2013);
[h]= not possible to derive as passing was an inclusion criteria for entering this study;
[i]= Not assessed in patients with ABI.

Incremental validity of the GET over already established PVTs

The incremental validity of the GET over the TOMM and DCT was explored in partial ROC analyses (pROC; see Figure 3). In this analysis, patients who failed either of the tests were included, resulting in 192 simulators and 55 patients with ABI with genuine cognitive dysfunctions. Regressing the DCT on the GET-ABI index revealed a significant model: $F(1, 244) = 79.05$, $p < .001$, with 24.5% explained variance. pROC analysis on the GET-ABI index to identify feigned cognitive dysfunction independently from the contribution of the DCT indicated good diagnostic accuracy: $pAUC = .849$, $SE = .025$, $CI = .800; .899$, and $p < .001$. Furthermore, regressing the TOMM on the GET-ABI scores also revealed a significant model, $F(2, 236) = 157.93$, $p < .001$, with explained variance reaching 57.2%. The pROC demonstrated that the GET-ABI index did not add significant diagnostic accuracy of

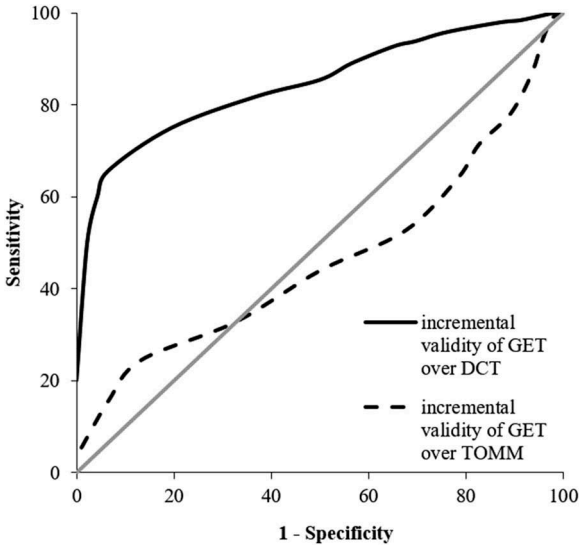

Figure 3. ROC curve depicting diagnostic accuracy of the GET-ABI index in identifying feigned cognitive dysfunction (n = 192) relative to genuine cognitive dysfunction (n = 55) after partialing out DCT or TOMM.

feigned cognitive dysfunction to the TOMM: $pAUC$ = .470, SE = .043, CI = .387; .553, and p < .526.

Simulation strategy use and face validity of PVTs

The majority of participants reported that they tried to feign cognitive dysfunction in the domain of executive functions (n = 135, 70.3%), attention (n = 102, 53.1%), speed of responses (n = 40, 20.8%) and memory (n = 34, 17.7%). With regard to these four simulation strategies, the four PVTs tests used in this study did not vary considerably in terms of identifying simulators across different cognitive domains (Table 6). The TOMM performed best, recognizing between 91.2% and 98.5% of participants applying a specific simulation strategy. The GET recognized between 82.4% and 89.2% of participants employing a specific simulation strategy, while the b Test recognized between 55.9% and 66.7%. The DCT, showing the lowest diagnostic accuracy, identified only between 26.5% and 33.3% of the simulators that employed specific strategies. Instructed simulators (n = 192) did not vary greatly in terms of which validity test they recognized as a validity test. If test takers are told that there are PVTs in the battery, they are most likely to pick the DCT as a validity test (n = 76; 39.6%), followed by the TOMM (n = 69; 35.9%), GET (n = 55; 28.6%), and b Test (n = 51; 26.6%).

Table 6. Diagnostic accuracy of various validity tests to detect feigning (N(%)) per simulation strategy used: speed, inattention, memory, executive functions.

	Speed (n = 40)	Inattention (n = 102)	Memory (n = 34)	Executive functions (n = 135)
GET-ABI index	34 (85.0)	91 (89.2)	28 (82.4)	117 (86.7)
TOMM	39 (97.5)	100 (98)	31 (91.2)	133 (98.5)
DCT	12 (30)	29 (28.4)	9 (26.5)	45 (33.3)
b Test	25 (62.5)	68 (66.7)	19 (55.9)	86 (63.7)

GET = Groningen Effort Test; DCT = Dot Counting Test; TOMM = Test Of Memory Malingering.

Discussion

Neuropsychological assessment using routine measures of cognition revealed that a sizable proportion of patients with ABI had impairments in attention (15–35%), visuoconstructive abilities (19–29%), and word fluency (33–48%). While cognitive impairments were observed in several aspects of cognition, these impairments did not appear to affect GET performance detrimentally. In fact, an inspection of test performance showed that the GET was relatively easy to perform for both patients with ABI and healthy controls, as demonstrated by low error rates in both groups (on average about 3–7 errors out of 94 items). Nevertheless, inferential statistics of group differences on GET performance reached significance levels, indicating effects of medium size. More specifically, patients with ABI had significantly longer response times than controls, while making significantly more mistakes. Instructed simulators on the other hand had significantly shorter response times (small effect) but committed significantly more errors (large effect) than patients with ABI. This might indicate a trade-off of speed for accuracy for instructed simulators.

With regard to the detection of feigned cognitive dysfunction, the TOMM was best in detecting feigning (sensitivity = 98.4%), closely followed by the GET which also showed excellent overall classification accuracy (AUC = .949; sensitivity of 88.5% at a specificity of 90%). The b Test and DCT showed an inferior performance to the GET and TOMM, with sensitivity levels of 70.3% and 39.1%, respectively, based on cutoffs as derived from the test manuals. Higher rates of sensitivity (80.7% for the DCT and 86.5% for the b Test) were reached based on the more recently suggested cutoffs. However, it can be argued that the traditional, lenient cutoffs may be more applicable for this sample of patients with brain injury, given the more impaired status of this sample relative to the typical heterogeneous neuropsychology clinic population. This is supported by the high failure rate of the DCT in the patient sample based on the revised cutoff (26%) compared to the traditional cutoff (7%). While the GET was shown to have good incremental validity over the DCT ($pAUC$ = .849), no incremental validity was achieved over the TOMM ($pAUC$ = .470). It must be noted that the classification accuracy of the GET-ABI index may be slightly overestimated, as the derivation of

the index and its evaluation was performed on the same sample. Nevertheless, the previously established GET-ADHD index also performed well in this context. Of note, the results are based on a study using a simulation design, which may over-estimate levels of sensitivity, as simulation designs are known to be limited in regard to the representation of real-life feigning. Given the results still hold in studies using *known-groups* methodology, it may be concluded that the GET is a useful instrument in distinguishing genuine from feigned cognitive dysfunction after ABI, with comparable classification rates to the TOMM, and comparable to superior accuracy (depending on the cutoff used) compared to the b Test and the DCT. It can further be concluded that the GET has a broad scope of application, and may, given further clinical validation, be a useful tool for the assessment of the performance validity of patients, not only in a psychiatric (Fuermaier et al., 2016) but also a neurological setting. Optimal application of the GET in various settings requires further validation in clinical practice using different research methodology, as well as the development of disorder-specific algorithms and cut-offs. This is suggested for the majority of already existing PVTs, and is also indicated for the GET since the present data in comparison to previous work using the GET (Fuermaier et al., 2016) show that the diagnostic accuracy of a single cutoff on a predefined variable varies per diagnostic group.

Since the validity of PVTs may be compromised when used repeatedly in the reassessment of a client, it is advantageous to have access to alternative measures with high diagnostic value (Carone, Iverson, & Bush, 2010). In this context, the GET can be a valuable tool. Furthermore, considering that about half of the instructed simulators of the present study had been coached to simulate cognitive dysfunctions, it can be assumed that the GET is robust against well-prepared individuals trying to feign cognitive dysfunction. The GET is also unlikely to be recognized by most participants as an validity measure due to its characteristics that resemble standard measures of attention. In the present study, fewer than a third of the simulators recognized the GET as a validity test, even though they were all aware that at least one of the tests administered was designed to assess performance validity. Moreover, as another advantage, the GET appears not only to be sensitive toward feigned attention disorders, but also toward other simulation strategies, such as executive dysfunction and memory impairment. Similar results were observed for other PVTs applied, showing that successful identification of feigning was largely independent of the simulation strategy used.

Limitations and future directions

The present study has to be seen in the context of several limitations. Simulation designs, by nature, can be criticized for limited external validity (Rogers & Cruise, 1998), as they may not present the simulating behavior of individuals in clinical practice. Different methodological approaches, such as *known-groups* comparisons (Rogers, Harrell, & Liff, 1993), should thus be used in future studies to complement the present findings. To ensure quality of the present data and increase the external validity of this study, strict exclusion criteria were applied in the selection of participants. Healthy controls and patients with ABI had to pass multiple established validity tests in order to ensure credibility of their test performances. Furthermore, instructed simulators were excluded if they scored below the cutoff of the RDS, indicating noncredible performance prior to simulation. It was also verified that instructed simulators had read and understood the instructions before performing the PVTs. Moreover, instructed simulators were asked after completion of the assessment whether they had applied their best effort in feigning cognitive dysfunction, and participants with coached simulation conditions were warned not to overdo their simulation and not to make their problems too obvious. Finally, participants were further motivated by the chance of winning a tablet PC, because past research in the field has shown that monetary incentives influence performance on clinical instruments (Binder & Rohling, 1996; Binder & Willis, 1991). However, more solid recommendations for clinical use can be given if it is known how the GET performs in studies using *known-groups* methodology.

Furthermore, it must be noted that standard instructions of neuropsychological testing were altered by informing test takers that at least one test was designed to assess performance validity, which may have influenced the participants' test performance. For example, by telling test takers that there are PVTs in the battery, they may pick a specific test as a PVT, and then reign in their feigning on this measure. Assuming this instruction affected test performance, it can be argued that it may have resulted in a more conservative estimation of classification rates. One may also argue that the comparison of classification rates across tests was still largely unaffected, given the rather similar rates with which the tests were recognized as a PVT.

Another caveat was that healthy individuals and instructed simulators were students and may therefore not be representative of the patient group (e.g., with regard to age, gender, and education). However, we do not regard this as a substantial problem for the present

study, as in general validity tests have been found to be largely independent of variables such as age and education, because only at extremes of these variables may distorted test results occur (Green & Flaro, 2003; Heilbronner et al., 2009). This may be one of the reasons why most established validity tests do not have age- or education-based norms but provide normative data based on diagnostic groups.

Furthermore, the composition of the patient sample and their disease severity may not be fully applicable to this context. As this study made use of a mixed ABI patient group consisting of patients with stroke or TBI, future research should aim for larger sample sizes to adequately differentiate patients and simulators according to the etiology of their condition. Furthermore, past research in this field has mainly studied patients with mild cases of ABI, which corresponds to the finding that individuals feigning cognitive deficits aiming for a financial payoff in personal injury litigation cases usually suffer from mild cognitive impairments (Reynolds & Horton, 2012). By including cases with more severe ABI in the present study, we showed that the GET can even be performed by more severely cognitively impaired patients, which may lead to an underestimation of the differences between patients with mild/moderate ABI and instructed stimulators.

Finally, coaching instructions may not be comparable to how individuals would prepare themselves before attempting to simulate cognitive dysfunction after ABI in real life. Past research has shown that people attempting to feign cognitive dysfunction tend to exaggerate deficits because of an underestimation of the functions retained and an overestimation of the test difficulty (Iverson & Franzen, 1998; Mittenberg, Azrin, Millsaps, & Heilbronner, 1993). Other research has shown that coaching participants with regard to simulation can attenuate this effect (Rose et al., 1998). Even though Suhr and Gunstad (2000) have demonstrated that cautionary remarks in the coaching instructions also lead to better malingering performances, the depth and detail of simulation strategies provided in the present context may not be representative of the preparation of an individual actually trying to feign cognitive dysfunction in real life.

Conclusion

The present study employed a simulation design to determine the utility of the GET in differentiating credible from noncredible cognitive dysfunction after ABI, i.e., stroke or TBI. A disorder-specific algorithm and cutoffs were derived which showed the GET to have excellent overall classification accuracy, with diagnostic accuracy levels close to the TOMM, and superior to the b Test and DCT. Furthermore, high diagnostic accuracy was stable and established independently from simulation strategies applied by simulators. Finally, face validity of the GET as a validity measure may be low, with only a minority of participants recognizing the GET as a validity measure. In conclusion, results of this simulation design support the GET's utility, in particular when applying disorder-specific algorithms, and thus suggests a broad field of application including patients with neurological conditions. The GET may complement neuropsychological assessments of patients with ABI, but needs further validation in studies using different research methodology, such as *known-groups* designs, before more solid clinical recommendations can be given.

Acknowledgments

We thank all research assistants involved in this project for their support in data collection and processing.

Disclosure statement

The authors A.B.M. Fuermaier, O. Tucha, S. Aschenbrenner, J. Koerts, and L. Tucha, have contracts with Schuhfried GmbH for the development of neuropsychological assessment tools.

Data availability statement

The data that support the findings of this study are available from the corresponding author, ABMF, upon reasonable request.

ORCID

Anselm B. M. Fuermaier http://orcid.org/0000-0002-2331-0840
Oliver Tucha http://orcid.org/0000-0001-8427-5279
Janneke Koerts http://orcid.org/0000-0002-2317-0171

References

Arnett, P., Hammeke, T., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-Item Test in neurological patients and dissimulators. *The Clinical Neuropsychologist*, *9*(1), 17–26.

Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *Regensburger Wortflüssigkeits-Test (RWT)*. Göttingen: Hogrefe.

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry*, *153*(1), 7.

Binder, L. M., & Willis, S. C. (1991). Assesment of motivation after financially compensable minor head trauma.

Psychological Assessment: A Journal of Consulting and Clinical Psychology, *3*(2), 175–181.
Binks, P. G., Gouvier, W. D., & Waters, W. F. (1997). Malingering detection with the dot counting test. *Archives of Clinical Neuropsychology*, *12*(1), 41–46.
Boone, K., Lu, P., & Herzberg, D. (2002a). *The dot counting test*. Los Angeles, CA: Western Psychological Service.
Boone, K., Lu, P., & Herzberg, D. (2002b). *The b Test*. Los Angeles, CA: Western Psychological Service.
Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. (K. B. Boone, Ed.). New York, NY: Guilford Press. Retrieved from http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2007-12979-000&site=ehost-live&scope=site
Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., … Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426.
Carone, D. A., Iverson, G. L., & Bush, S. S. (2010). A model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist*, *24*(5), 759–778.
Crawford, J. R., Parker, D. M., & Besson, J. A. (1988). Estimation of premorbid intelligence in organic conditions. *The British Journal of Psychiatry*, *153*, 178–181.
Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six european countries. *Archives of Clinical Neuropsychology*, *28*(8), 771–783.
Erdal, K. (2004). The effects of motivation, coaching, and knowledge of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology*, *19*(1), 73–88.
Fuermaier, A. B., Tucha, O., Koerts, J., Grabski, M., Lange, K. W., Weisbrod, M., … Tucha, L. (2016). The development of an embedded figures test for the detection of feigned attention deficit hyperactivity disorder in adulthood. *PloS One*, *11*(10), e0164297.
Fuermaier, A. B. M., Tucha, L., Koerts, J., Aschenbrenner, S., & Tucha, O. (2017). *Vienna Test System (VTS): The Groningen Effort Test (GET) for the assessment of noncredible cognitive performance in the assessment of adult ADHD*. Vienna, Austria: Schuhfried.
Green, P. (2005a). *Medical symptom validity test*. Edmonton: Green's Publishing.
Green, P. (2005b). *Nonverbal medical symptom validity test*. Edmonton: Green's Publishing.
Green, P., Allen, L., & Astner, K. (1996). *Manual for computerised Word Memory Test*. Durham, NV: CogniSyst.
Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, *9*(3), 189–207.
Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesic measures with a large clinical sample. *Psychological assessment*, *6*, 218–224.
Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., … Ciota, M. A. (2007). Malingering in toxic exposure: Classification accuracy of reliable digit span and WAIS-III digit span scaled scores. *Assessment*, *14*(1), 12–21.
Haines, M. E., & Norris, M. P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits. *The Clinical Neuropsychologist*, *15*(2), 171–182.
Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129.
Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS digit span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment*, *12*(4), 429–444.
Iverson, G. L., & Franzen, M. D. (1998). Detecting malingered memory deficits with the Recognition Memory Test. *Brain Injury*, *12*(4), 275–282.
Jennekens, N., de Casterlé, B. D., & Dobbels, F. (2010). A systematic review of care needs of people with traumatic brain injury (TBI) on a cognitive, emotional and behavioural level. *Journal of Clinical Nursing*, *19*(9–10), 1198–1206.
Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.
Lehrl, S. (1995). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Balingen: Spitta-Verlag.
Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple-Choice Vocabulary-Test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, *91*(5), 335–345.
Lezak, M. D. (2004). *Neuropsychological assessment*. USA: Oxford University Press.
McCaul, C., Boone, K. B., Ermshar, A., Cottingham, M., Victor, T. L., Ziegler, E., … Wright, M. (2018). Cross-validation of the Dot Counting Test in a large sample of credible and non-credible patients referred for neuropsychological testing. *The Clinical Neuropsychologist*, *32*(6), 1054–1067.
Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale—Revised. *Psychological Assessment*, *5*(1), 34–40.
Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exeggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102.
Petermann, F. (2012). *Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV). German version*. Frankfurt: Pearson Assessment.
Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*(1), 10–20.
Reynolds, C. R., & Horton, A. J. (2012). *Detection of malingering during head injury litigation* (2nd ed.). New York, NY: Springer Science + Business Media. doi:10.1007/978-1-4614-0442-2
Roberson, C. J., Boone, K. B., Goldberg, H., Miora, D., Cottingham, M., Victor, T., … Wright, M. (2013). Cross validation of the b test in a large known groups sample. *The Clinical Neuropsychologist*, *27*(3), 495–508.

Rogers, R. (2008). *Clinical assessment of malingering and deception*. New York, NY: Guilford Press.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*(3), 273–285.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment—A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*(3), 255–274.

Rose, F. E., Hall, S., Szalda-Petree, A. D., & Bach, P. J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology*, *13*(4), 349–363.

Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, *15*(5), 415–424.

Sun, J. H., Tan, L., & Yu, J. T. (2014). Post-stroke cognitive impairment: Epidemiology, mechanisms and management. *Annals of Translational Medicine*, *2*(8), 80.

Suslow, T. (2009). Estimating verbal intelligence in unipolar depression: Comparison of word definition and word recognition. *Nordic Journal of Psychiatry*, *63*(2), 120–123.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, *16*(4), 495–505.

Tombaugh, T. N. (1996). *Test of memory malingering*. Toronto, Canada: Multi-Health Systems.

Zimmermann, P., & Fimm, B. (2010). *TAP: Test for attentional performance – Version 2.2*. Herzogenrath, Germany: Psytest.